https://doi.org/10.2352/CIC.2024.32.1.16

Exploring the efficiency of end-to-end vs. separate sequencing of DNN-based AWB and denoising in-camera processing pipeline

Shuwei Yue, Minchen Wei*; The Hong Kong Polytechnic University, Hong Kong
minchen.wei@polyu.edu.hk

Abstract

In modern image signal processors (ISPs), many modules have adopted deep neural networks (DNNs). This study explores whether a single DNN can effectively replace both the auto white balance (AWB) and denoising modules, or if they should be processed separately. Our experiment results suggest that performing AWB and denoising individually can produce better performance than an end-to-end approach. Moreover, processing denoising before AWB leads to a significant improvement, with an increase of nearly ***6 dB*** *in PSNR and* ***30%*** *reduction in mean angular error (MAE). These findings suggest that careful consideration of the processing order in ISP pipelines can lead to substantial enhancements in image quality.*

Introduction

Image signal processors (ISPs) are crucial in modern digital imaging systems, converting raw sensor data into high-quality images. With the development of deep learning, many ISP modules now incorporate deep neural networks (DNNs) to enhance performance. Previous research has demonstrated the effectiveness of DNNs in individual ISP tasks [1–4, 6]. Furthermore, combining modules such as denoising and demosaicing [7], denoising and tone mapping [5], denoising and super-resolution [10], denoising, demosaicing, and super-resolution [8, 9], and even the entire ISP [12] was also found to reduce accumulated errors and computational complexity.

However, AWB and denoising, being the two early steps in the ISP pipeline, have not been thoroughly studied about whether they can be combined into an end-to-end module or not. The interaction between AWB and denoising is particularly significant, as noise can affect the accuracy of AWB and incorrect white balance can change the noise distribution and then, affect the effectiveness of denoising. To address this problem, we aim to answer two specific questions in this study: 1) Can AWB and denoising be effectively replaced by a single DNN? 2) If it is desired to perform them individually, which order can produce better image quality: AWB followed by denoising (AWB & denoising) or denoising followed by AWB (denoising & AWB)?

To explore these questions, we re-processed the benchmark dataset [11] and conducted thorough experiments. Our results show that performing AWB and denoising individually can outperform the end-to-end DNN approach. Specifically, performing denoising before AWB improves image quality by 6 dB in PSNR and 30% in MAE compared to the reverse order, as is shown in Figure 1. These findings highlight the importance of sequencing order in ISP pipelines and offer valuable insights for future DNN-based ISP design and optimization.

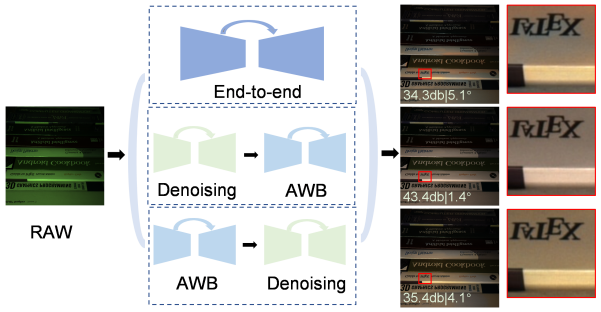

Figure 1. *Illustration of the three approches investigated in this study: an end-to-end and two individual approaches (denoising before AWB, and AWB before denoising) using the re-processed SIDD dataset [11].*

Related work

We discuss DNN or AI-based AWB, denoising, ISP modules, and the importance of the order of the ISP modules here.

AI AWB AI-based AWB methods like FC4 [13], DMCC [17], and LSMIU [16] have outperformed traditional methods such as grey-world [14], shades of grey [18], and PCA-based [15] across various scenes [1]. We use LSMIU, a UNet-based architecture, as our baseline.

AI Denoising Traditional denoising methods such as median filtering [20], bilateral filtering [21], and non-local means (NLM) [22] have limited performance under extreme dark conditions. AI denoising methods, such as those using DarkU [23], have shown superior performance, particularly in the RAW domain [19, 24, 25]. We use DarkU as our baseline due to its generalizability and effectiveness.

AI ISP In addition to using a DNN to replace single modules, researchers have explored using a single DNN to replace multiple modules, such as denoising and demosaicking [7, 8], denoising and super-resolution [10], and even three modules like denoising, demosaicking, and super-resolution [9], or the entire ISP pipeline [12]. Other researchers have proposed multi-stage approaches to replace the whole ISP, such as two-stage methods [27] for image restoration and enhancement, and three-stage [28] methods for handling noise, color, and brightness. Here, we investigate whether AWB and denoising modules can be replaced by a single end-to-end DNN to mitigate the potential accumulate error, which is a unique approach.

Order of modules in ISP Most papers focus on the order of denoising and demosaicking. The advantage of denoising then de-

mosaicing [2, 30] is that the Poisson noise can be transformed to white Gaussian noise using an Anscombe transform. However, a comprehensive study [26] found that demosaicing and then denoising is generally better. To the best of our knowledge, we are the first to explore the sequencing of DNN-based AWB and denoising, which is crucial for the design of modern ISPs.

Method

Problem definition This study aims to explore the effectiveness of using an end-to-end DNN for AWB and denoising, and the optimal order of performing AWB and denoising individually. We compare the performance of the following approaches:

- **End-to-end approach:** Applying a single DNN to perform both AWB and denoising simultaneously.
- **Separate sequencing:** Applying AWB and denoising individually, with two orders:
  - AWB & denoising
  - Denoising & AWB

Evaluation metrics We use two metrics to evaluate the quality of the final processed images: PSNR (Peak Signal-to-Noise Ratio) and MAE (Mean Angular Error) to calculate the noise and color error, respectively. For an RGB image, PSNR is calculated by taking the mean squared error(*mse*) across all channels and pixels with $\text{PSNR} = 10 \cdot \log_{10}\left(\frac{1.0^2}{mse}\right)$. MAE measures the average angular error between the predicted image $\hat{\mathbf{y}}_i$ and the ground truth $\mathbf{y}_i$ for all the pixels N in an image: $\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{180}{\pi}\arccos\left(\mathbf{y}_i \cdot \hat{\mathbf{y}}_i\right)$.

These metrics allow us to comprehensively evaluate the effectiveness of the final image quality after AWB and denoising or vice versa.

Model selection For the end-to-end strategy, we employ two architectures: DarkU from [23] and UNet++ [31], which enhances feature fusion through multi-level skip connections and a multi-path structure compared to DarkU. While there are more advanced DNN models available, our goal is to compare the effectiveness of the end-to-end and individual processing strategies. To ensure a fair comparison, we need to control variables; thus, if a more advanced model is used for the end-to-end processing, the same model should also be used for the individual processing. Given that AI AWB is well-represented by LSMIU and AI denoising by DarkU, we chose DarkU and UNet++ for our experiments. Such a choice allows us to maintain comparable architectures and parameters between the two strategies, ensuring a reasonable and controlled comparison.

Dataset selection and preprocessing We utilize the SIDD (Smartphone Image Denoising Dataset) [11], which is specifically designed for real image denoising tasks. SIDD provides a comprehensive set of noisy and clean image pairs captured under various lighting conditions and different ISO settings, with the higher ISO settings resulting in stronger noise. This makes it an ideal dataset for evaluating the performance of our models in real-world scenarios. The dataset includes RAW-noisy and RAW-clean image formats, as well as metadata such as the white point gains, Bayer pattern, and color correction matrix.

We preprocess the raw format to obtain the demosaiced RGB-noisy image through the bilinear interpolation algorithm [32], the corresponding white points (AWB ground truth), clean RGB without AWB, and the final clean RGB with AWB for different experiments. To verify the robustness of our strategy, we manually divide the images into the low (5-200) and high (201-1600) ISO categories. After removing obviously color-incorrect images, we included 122 Low-ISO images and 116 High-ISO images in total. We randomly split the images into 70% for training, 20% for validation, and 10% for testing. All performance evaluations are conducted on the testing set. The details of the experiments and data usage are listed in Table 1.

Table 1. Different approaches and corresponding data usage.

Experiments\Data	Noisy-RGB	Clean-RGB (without wb)	Clean-RGB (with wb)	White points
AWB only	✓	✗	✗	✓
Denoising only	✓	✓	✗	✗
AWB & denoising	✓	✗	✓	✓
denoising & AWB	✓	✓	✗	✓
End2end	✓	✗	✓	✗

Experiments and results

Settings Our method is implemented using the PyTorch library and trained on NVIDIA RTX4090 GPUs. The training images are center-cropped to a size of 1024x1024 as input and output. We use the Adam optimizer [29] with a cosine decay strategy, a maximum of 500 epochs, a batch size of 16, and a learning rate of 4e-4.

Results We evaluated the performance on the re-processed SIDD dataset under two ISO ranges: Low-ISO and High-ISO. For the end-to-end approach, the performance metrics, shown in Table 2, indicate lower PSNR and higher MAE. In contrast, the individual strategies, specifically denoising & AWB and AWB & denoising orders, significantly outperformed the end-to-end approach. The denoising & AWB order was found to be far superior to the reverse order, with a nearly 6 dB improvement in PSNR and a 30% enhancement in MAE. Qualitative results as shown in Figure 2.

Discussion

Why denoising & AWB works better than AWB &denoising The experiment results indicate that performing denoising & AWB significantly improves image quality. One explanation for such a finding is that AWB algorithms are sensitive to noise. When noise is present, the AWB algorithm may incorrectly estimate the white point, leading to poor color correction, and further accumulating the color error to influence the subsequent denoising processing. This is supported by an ablation study, in which only using AWB under the two ISO settings showed that MAE and illumination angular error are drastically reduced under high ISO conditions, as shown in Table 3 In other words, by removing noise first, the AWB algorithm can result in more accurate color balance and better overall image quality.

Do Non-DNN and DNN methods perform the same for sequencing As we aim to verify the importance of the order in DNN-based methods, one may ask if this still applies to non-DNN-based methods. We conducted an ablation study using traditional AWB methods such as Grey-world [14], PCA-based [15](p=3%), and SoG [18]($p = 3$), as well as denoising methods such as median filter ($f = 3$) and bilateral filter ($\sigma_{\text{color}} =$

Table 2. Comparison between the end-to-end and individual strategies (denoising & AWB, AWB & denoising) across different ISO levels, in terms of PSNR (higher is better) and MAE (lower is better).

Strategy\ Noisy level		Low-ISO (5-200)		High-ISO (201-1600)	
		PSNR↑	MAE↓	PSNR↑	MAE↓
End-to-end	DarkU [23]	29.5	4.2	28.8	5.1
	UNet++ [31]	28.5	5.1	27.4	6.1
Individual	denoising & AWB	**36.0**	**2.8**	**33.4**	**3.0**
	AWB & denoising	30.5	4.1	28.8	4.2

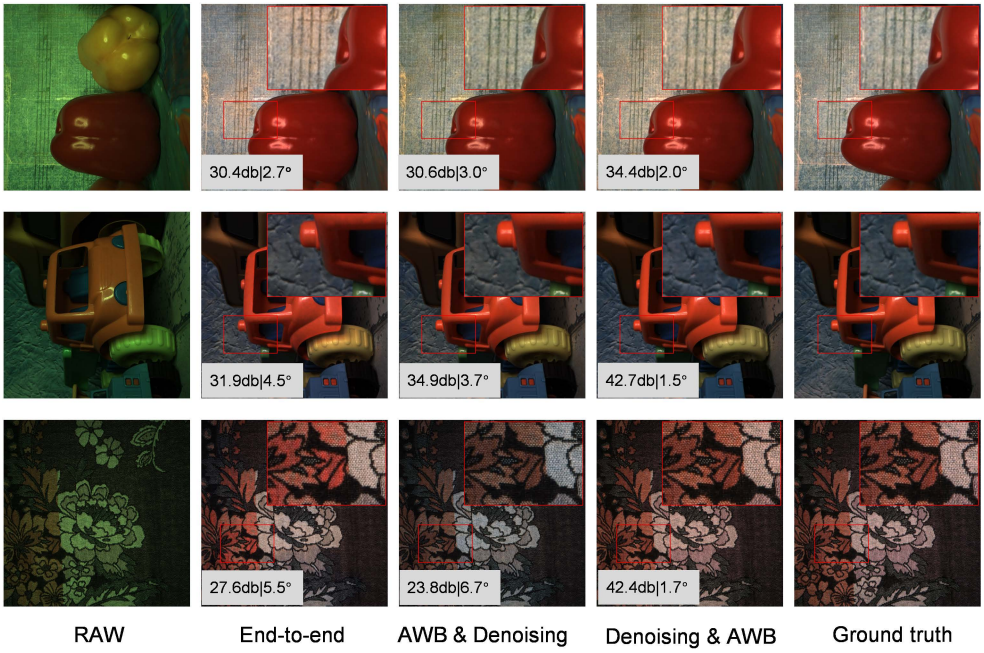

Figure 2. *Qualitative examples of images processed using different approaches: End-to-end, AWB before Denoising, Denoising before AWB, and Ground Truth. The values shown are the PSNR (dB) and MAE (degrees).*

Table 3. MAE values for only performing AWB under different ISO conditions.

	Low-ISO	High-ISO
AWB only	3.4	5.8

75, $\sigma_{\text{space}} = 75$). The results, as shown in Table 4, indicate that the order of the two modules is only important for DNN-based methods. Another interesting finding is that the grey-world method is not sensitive to noise levels, while PCA-based and SoG are very sensitive. This is reasonable as the latter methods use bright pixels, which can be severely influenced by noise levels.

Conclusion

In this study, we investigated the efficiency of using a DNN-based end-to-end strategy to replace the AWB and denoising functions in ISP, and concluded that this approach is not effective. Additionally, we examined the optimal sequencing for AWB and denoising when processed separately. Our findings show that performing denoising before AWB (denoising & AWB) significantly improves image quality, with an increase of nearly 6 dB in PSNR and a 30% enhancement in MAE compared to the reverse order. These findings underscore the critical importance of sequencing design in DNN-based ISP pipelines.

References

[1] S. Yue and M. Wei, Color constancy from a pure color view, J. Opt. Soc. Am. A., 40, 602 (2023).

[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, Beyond a gaussian denoisingr: Residual learning of deep cnn for image denoising, IEEE Trans. Image Process., 26, pg. 3142 (2017).

[3] N. S. Syu, Y. S. Chen, and Y. Y. Chuang, Learning deep convolutional networks for demosaicing, arXiv preprint arXiv:1802.03769 (2018).

[4] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, Deep tone mapping operator for high dynamic range images, IEEE Trans. Image Process., 29, pg. 1285 (2019).

[5] L. Hu, H. Chen, and J. P. Allebach, Joint multi-scale tone mapping and denoising for HDR image enhancement, In Proc. *WACV*, pg. 729 (2022).

[6] M. Ehrlich and L. S. Davis, Deep residual learning in the jpeg transform domain, In Proc. *ICCV*, pg. 3484 (2019).

[7] T. Ehret, A. Davy, P. Arias, and G. Facciolo, Joint demosaicking and denoising by fine-tuning of bursts of raw images, In Proc. *ICCV*, pg. 8868 (2019).

Table 4. Comparison of Non-DNN and DNN-based methods for different approaches.

	AWB	Denoising	AWB & denoising (PSNR/MAE)	denoising & AWB (PSNR/MAE)
Non-DNN	Grey-world [14]	Median-filter	24.3/10.4	24.5/10.0
		Bilateral-filter	24.9/9.7	24.8/9.5
	PCA-based [15]	Median-filter	15.4/33.5	15.1/32.1
		Bilateral-filter	15.1/33.3	15.2/31.8
	Sog [18]	Median-filter	18.4/11.3	18.5/11.0
		Bilateral-filter	18.5/10.7	18.6/10.5
DNN-based	LSMIU [16]	DarkU [23]	30.5/4.1	36.0/2.8

[8] F. Kokkinos and S. Lefkimmiatis, Iterative joint image demosaicking and denoising using a residual denoising network, IEEE Trans. Image Process., 28, pg. 4177 (2019).

[9] W. Xing and K. Egiazarian, End-to-end learning for joint image demosaicing, denoising and super-resolution, In Proc. *CVPR*, pg. 3507 (2021).

[10] J. Xie, R. S. Feris, S.-S. Yu, and M.-T. Sun, Joint super resolution and denoising from a single depth image, IEEE Trans. Multimedia., 17, pg. 1525 (2015).

[11] A. Abdelhamed, S. Lin, and M. S. Brown, A high-quality denoising dataset for smartphone cameras, In Proc. *CVPR*, pg. 1692 (2018).

[12] A. Ignatov, L. Van Gool, and R. Timofte, Replacing mobile camera isp with a single deep learning model, In Proc. *CVPRW*, pg. 536 (2020).

[13] Y. Hu, B. Wang, and S. Lin, Fc4: Fully convolutional color constancy with confidence-weighted pooling, In Proc. *CVPR*, pg. 4085 (2017).

[14] G. Buchsbaum, A spatial processor model for object colour perception., J. Franklin Inst., 310, 1 (1980).

[15] D. Cheng, D. K. Prasad, and M. S. Brown, Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution, J. Opt. Soc. Am. A,., 31, pg. 1049 (2014).

[16] D. Kim, J. Kim, S. Nam, D. Lee, Y. Lee, N. Kang, H. E. Lee, B. Yoo, J. J. Han, and S. J. Kim, Large scale multi-illuminant (LSMI) dataset for developing white balance algorithm under mixed illumination, In Proc. *CVPR*, pg. 2410 (2021).

[17] S. Yue and M. Wei, Effective cross-sensor color constancy using a dual-mapping strategy, J. Opt. Soc. Am. A., 41, pg. 329 (2024).

[18] G. D. Finlayson and E. Trezzi, Shades of gray and colour constancy, Color and Imaging Conference, pg. 37 (2004).

[19] X. Jin, W. Xiao, H. Han, C. Guo, R. Zhang, X. Liu, and C. Li, Lighting every darkness in two pairs: A calibration-free pipeline for raw denoising, In Proc. *ICCV*, pg. 13275 (2023).

[20] G. Qiu, An improved recursive median filtering scheme for image processing, IEEE Trans. Image Process., 5, pg. 646 (1996).

[21] M. Elad, On the origin of the bilateral filter and ways to improve it, IEEE Trans. Image Process., 11, pg. 1141 (2002).

[22] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, Adaptive non-local means denoising of MR images with spatially varying noise levels, J. Magn. Reson. Imaging, 31, pg. 192 (2010).

[23] C. Chen, Q. Chen, J. Xu, and V. Koltun, Learning to see in the dark, In Proc. *CVPR*, pg. 3291 (2018).

[24] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, Noise2Noise: Learning image restoration without clean data, arXiv preprint arXiv:1803.04189 (2018).

[25] K. Wei, Y. Fu, J. Yang, and H. Huang, A physics-based noise formation model for extreme low-light raw denoising, In Proc. *CVPR*, pg. 2758 (2020).

[26] Q. Jin, G. Facciolo, and J.-M. Morel, A review of an old dilemma: Demosaicking first, or denoising first?, In Proc. *CVPRW*, pg. 514 (2020).

[27] Z. Liang, J. Cai, Z. Cao, and L. Zhang, Cameranet: A two-stage framework for effective camera isp learning, IEEE Trans. Image Process., 30, pg. 2248 (2021).

[28] S. Liu, C. Feng, X. Wang, H. Wang, R. Zhu, Y. Li, and L. Lei, Deepflexisp: A three-stage framework for night photography rendering, In Proc. *CVPR*, pg. 1211 (2022).

[29] D. Kingma, A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[30] O. Kalevo and H. Rantanen, Noise reduction techniques for Bayer-matrix images, In Proc. *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III*, pg. 348 (2002).

[31] Z. Zhou, M. R. Siddiquee, N. Tajbakhsh, and J. Liang, Unet++: A nested UNet architecture for medical image segmentation, In Proc. *DLMIA*, pg. 3 (2018).

[32] H. S. Malvar, L. He, and R. Cutler, High-quality linear interpolation for demosaicing of Bayer-patterned color images, In Proc. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pg. iii–485 (2004).